WHAT IS CLAIMED IS:

1. A method of mapping a set of input patterns to an m-dimensional space, comprising the steps of:

(a) selecting k patterns from said set of input patterns to form a subset of patterns $\{\mathbf{p}_i, i = 1, \ldots, k\}$;

(b) determining at least some pairwise relationships between at least some of the patterns in said subset of patterns $\{\mathbf{p}_i\}$;

(c) mapping the patterns $\{\mathbf{p}_i\}$ into a set of images in an m-dimensional space ($\mathbf{p}_i \to \mathbf{y}_i$, $i = 1, 2, \ldots k$, $\mathbf{y}_i \in R^m$) so that at least some of the pairwise distances between at least some of the images $\{\mathbf{y}_i\}$ are representative of the relationships of the respective patterns $\{\mathbf{p}_i\}$;

(d) determining a set of n attributes for each pattern in said subset of patterns $\{\mathbf{p}_i\}$, $\{\mathbf{x}_i, i = 1, 2, \ldots k, \mathbf{x}_i \in R^n\}$;

(e) forming a training set $T = \{(\mathbf{x}_i, \mathbf{y}_i), i = 1, 2, \ldots k\}$;

(f) using a supervised machine learning technique to determine a mapping function based on the training set T; and

(g) using said mapping function determined in step (f) to map additional patterns.

2. The method of claim 1, wherein said mapping function is encoded in at least one neural network.

3. The method of claim 1, wherein step (f) further comprises the steps of:

(i) determining c n-dimensional reference points, $\{\mathbf{c}_i, i = 1, 2, \ldots c, \mathbf{c}_i \in R^n\}$;

(ii) partitioning T into c disjoint clusters C_j based on a distance function d, $\{C_j = \{(\mathbf{x}_i, \mathbf{y}_i): d(\mathbf{x}_i, \mathbf{c}_j) \leq d(\mathbf{x}_i, \mathbf{c}_k)$ for all $k \neq j$; $j = 1, 2, \ldots c$; $i = 1, 2, \ldots k\}$; and

(iii) using a supervised machine learning technique to determine c independent mapping functions $\{f_i^L, i = 1, 2, \ldots c\}$, based on the respective subsets C_i of the training set T.

4. The method of claim 3, wherein each said mapping function is encoded in at least one neural network.

5. The method of claim 3, wherein step (g) further comprises the steps of:

(i) for an additional input pattern **p**, determining a set of n attributes $\mathbf{x}$, $\mathbf{x} \in R^n$;

(ii) determining the distance of $\mathbf{x}$ from each said reference point $\{\mathbf{c}_i\}$;

(iii) identifying the reference point $\mathbf{c}_j$ closest to $\mathbf{x}$; and

(iv) mapping $\mathbf{x} \rightarrow \mathbf{y}$, $\mathbf{y} \in R^m$, using the mapping function f_j^L associated with the reference point $\mathbf{c}_j$ identified in step (iii).

6. The method of claim 5, wherein each said mapping function is encoded in at least one neural network.

7. The method of claim 3, wherein step (i) is performed using a clustering algorithm.

8. The method of claim 1, wherein step (f) further comprises the steps of:

(i) determining c m-dimensional reference points, $\{\mathbf{c}_i, i = 1, 2, \ldots c, \mathbf{c}_i \in R^m\}$;

(ii) partitioning T into c disjoint clusters C_j based on a distance function d, $\{C_j = \{(\mathbf{x}_i, \mathbf{y}_i): d(\mathbf{y}_i, \mathbf{c}_j) \leq d(\mathbf{y}_i, \mathbf{c}_k) \text{ for all } k \neq j; j = 1, 2, \ldots c; i = 1, 2, \ldots k\}\}$;

(iii) using a supervised machine learning technique to determine c independent local mapping functions $\{f_i^L, i = 1, 2, \ldots c\}$ based on the respective subsets C_i of the training set T; and

(iv) using a supervised machine learning technique to determine a global mapping function f^G, based on the entire training set T.

9. The method of claim 8, wherein each said mapping function is encoded in at least one neural network.

10. The method of claim 8, wherein step (g) further comprises the steps of:

(i) for an additional input pattern **p**, determining a set of n attributes $\mathbf{x}$, $\mathbf{x} \in R^n$;

(ii) mapping $\mathbf{x} \rightarrow \mathbf{y'}$, $\mathbf{y'} \in R^m$, using the global mapping function f^G;

(iii) determining the distance of $\mathbf{y'}$ to each reference point in $\{\mathbf{c}_i\}$;

(iv) identifying the reference point $\mathbf{c}_j$ closest to $\mathbf{y'}$; and

(v) mapping $\mathbf{x} \rightarrow \mathbf{y}$, $\mathbf{y} \in R^m$, using the local mapping function f_j^L associated with the reference point $\mathbf{c}_j$ identified in step (iv).

11. The method of claim 10, wherein each said mapping function is encoded in at least one neural network.

12. The method of claim 8, wherein step (i) is performed using a clustering algorithm.

13. The method of claim 1, wherein step (c) further comprises the steps of:

(i) selecting a subset of patterns from $\{\mathbf{p}_i\}$;

(ii) revising the positions of the images of said selected subset of patterns in the m-dimensional space based on the relationships

between said selected subset of patterns determined in step (b); and

(iii) repeating steps (i) and (ii) for additional subsets of patterns from $\{\mathbf{p}_i\}$.

14. The method of claim 1, wherein step (d) comprises the step of:

(i) determining a set of n attributes for each pattern in said subset of patterns $\{\mathbf{p}_i\}$, $\{\mathbf{x}_i, i = 1, 2, \ldots k, \mathbf{x}_i \in R^n\}$, wherein said attributes represent the relationships of each pattern in said subset of patterns with respect to n other reference patterns.

15. The method of claim 1, wherein step (b) comprises the step of receiving pairwise relationship data from a subject.

16. The method of claim 15, wherein step (b) comprises the steps of:

(1) randomly selecting two patterns from the plurality of patterns;

(2) presenting the two patterns to at least one subject; and

(3) receiving pairwise relationship data from said subject about the patterns.

17. The method of claim 16, further comprising the step of:

(4) repeating steps (b)(1) through (b)(3) for additional pairs of patterns.

18. The method of claim 17, wherein step (b) further comprises the step of receiving data about the subjects providing the pairwise relationship data.

19. The method of claim 1, wherein step (b) comprises the steps of:

(1) receiving pairwise relationship data via a communications path coupled to a computer; and

(2) storing the received pairwise relationship data in a memory.

20. The method of claim 1, wherein step (b) comprises the steps of:

(1) selecting a pair of patterns for similarity comparison;

(2) transmitting information about the selected pair of patterns to a remote computer system; and

(3) receiving pairwise relationship data about the selected pair of patterns from the remote computer system.

21. The method of claim 20, further comprising the step of:

(4) repeating steps (b)(1) through (b)(3) for additional pairs of compounds.

22. The method of claim 21, wherein step (b)(1) comprises the step of selecting the pair of patterns at random.

23. A computer program product comprising a computer useable medium having computer program logic recorded thereon for enabling a processor to obtain pairwise relationship data about the selected pair of patterns, said computer program logic comprising:

a selecting procedure that enables the processor to select a plurality of patterns from a database for similarity;

a transmitting procedure that enables a processor to transmit selected patterns to a remote computer for similarity comparison; and

a receiving procedure that enables the processor to receiving similarity data about patterns transmitted to the remote computer.

24. The computer program product of claim 23, further comprising:

a storing procedure that enables the processor to store received similarity data for subsequent retrieval.

25. The computer program product of claim 24, wherein said selecting procedure randomly selecting patterns from the database.

26. A system for mapping a set of input patterns to an *m*-dimensional space, comprising:

means for selecting *k* patterns from said set of input patterns to form a subset of patterns$\{\mathbf{p}_i, i = 1, \ldots, k\}$;

means for determining at least some pairwise relationships between at least some of the patterns in said subset of patterns $\{\mathbf{p}_i\}$;

means for mapping the patterns $\{\mathbf{p}_i\}$ into a set of images in an *m*-dimensional space ($\mathbf{p}_i \rightarrow \mathbf{y}_i, i = 1, 2, \ldots k, \mathbf{y}_i \in R^m$) so that at least some of the pairwise distances between at least some of the images $\{\mathbf{y}_i\}$ are representative of the relationships of the respective patterns $\{\mathbf{p}_i\}$;

means for determining a set of *n* attributes for each pattern in said subset of patterns $\{\mathbf{p}_i\}$, $\{\mathbf{x}_i, i = 1, 2, \ldots k, \mathbf{x}_i \in R^n\}$;

means for forming a training set $T = \{(\mathbf{x}_i, \mathbf{y}_i), i = 1, 2, \ldots k\}$;

means for using a supervised machine learning technique to determine a mapping function based on the training set T; and

means for using said mapping function determined in step (f) to map additional patterns.

27. A computer program product comprising a computer useable medium having computer program logic recorded thereon for enabling a processor to obtain map a set of input patters to an *m*-dimensional space, said computer program logic comprising:

a procedure that selects *k* patterns from said set of input patterns to form a subset of patterns$\{\mathbf{p}_i, i = 1, \ldots, k\}$;

a procedure that determines at least some pairwise relationships between at least some of the patterns in said subset of patterns $\{\mathbf{p}_i\}$;

a procedure that maps the patterns $\{\mathbf{p}_i\}$ into a set of images in an m-dimensional space ($\mathbf{p}_i \rightarrow \mathbf{y}_i$, $i = 1, 2, \ldots k$, $\mathbf{y}_i \in R^m$) so that at least some of the pairwise distances between at least some of the images $\{\mathbf{y}_i\}$ are representative of the relationships of the respective patterns $\{\mathbf{p}_i\}$;

a procedure that determines a set of n attributes for each pattern in said subset of patterns $\{\mathbf{p}_i\}$, $\{\mathbf{x}_i, i = 1, 2, \ldots k, \mathbf{x}_i \in R^n\}$;

a procedure that forms a training set $T = \{(\mathbf{x}_i, \mathbf{y}_i), i = 1, 2, \ldots k\}$;

a procedure that uses a supervised machine learning technique to determine a mapping function based on the training set T; and

a procedure that uses said mapping function determined in step (f) to map additional patterns.